

## Oponentní posudek diplomové práce Bc. Miroslava Minářčika

Název práce: „Spektroskopické a statistické metody pro detailní mapování vegetace v Krkonošském národním parku“

Diplomová práce Miroslava Minářčika se zabývá z pohledu magisterského stupně studia náročnějším tématem využití statistických metod korespondenční analýzy a parciální regrese (PLSR) pro klasifikaci hyperspektálních dat. Použité metody jsou aplikovány na modelovém území krkonošské tundry. Jako vstupní data byla použita šetření z terénu, tj. botanikem definované druhové složení na vybraných plochách reprezentujících osm zvolených kategorií vegetace a jejich spektrální projev měřený spektrometrem typu ASD FieldSpec, dále pak hyperspektrální snímky pořízené senzorem APEX. Pro ukázání rozdílu statického přístupu klasifikace vůči běžnějším metodám byla provedena také řízená klasifikace s využitím algoritmu podpůrných vektorů (SVM). Diplomová práce má standardní členění do sedmi kapitol. Jejich uspořádání je až na drobné výjimky uvedené níže logické. Cíle práce jsou jasně definované, volba statického přístupu vyhodnocení dat pro floristicky heterogenní oblast je dostatečně zdůvodněna.

První část literární rešerše popisuje základní principy a využití vícerozměrných statistických metod pro analýzu druhového složení vegetace (vztahy mezi výskytem jednotlivých druhů, vazby mezi stanovišti, popř. charakteristikami přírodního prostředí). Zvláštní pozornost je věnována dále používané metodě korespondenční analýzy DCA. K úvodní části rešerše mám následující komentář. Text vychází prakticky z jediného literárního zdroje Herben a Münbergová (2003). Veškerá vysvětlení jsou pouze slovní, bez uvedení jediného matematického vztahu charakterizujícího prováděné transformace s daty. Domnívám se, že matematické vyjádření s krátkým příkladem by usnadnilo porozumění výkladu metody i uváděných parametrů analýzy. Uvedené příklady na obr. 3 až 6 by zasloužily bližší vysvětlení – co jednotlivé symboly resp. písmena/číslíce vyjadřují, podle jakého pravidla se volí velikost segmentu v obr. 3 a jaký má tento parametr vliv na výsledek, z jakého důvodu se data po transformaci zobrazí zrcadlově v obr. 5? Proč nejsou překládány zaběhnuté či snadno přeložitelné termíny, např. na str. 19? Zabránilo by se tak i spojením typu „detrendování trendu“ (str. 17). Zavádějící je vysvětlení termínu vlastních čísel (str. 19), kde je na jednu stranu uvedeno, že jsou vyjádřeny v „absolutních hodnotách“ a přitom vyjadřují „míru separace ...“, kde maximální hodnota je 1“, což jasně evokuje relativní vyjádření. Ke zbývajícím částem rešerše nemám výhrady. Naopak kapitoly 2.4 až 2.6 uvádějící příklady prací využívajících podobné postupy považuji za velmi dobře zpracované, stejně jako následující kapitolu zpracovávající fyzicko-geografický a botanický popis zájmového území.

V kapitole data a metodika bych doporučovala obr. 13 rozšířit o příklad podrobnějšího zobrazení umístění dílčích studovaných ploch 1 m x 1 m (obrázek ukazuje pouze přehled ploch odpovídajících jednotlivým třídám). Předpokládám, že rozmístění a volba těchto ploch, stejně jako veškerá další data byly pro diplomovou práci dány a tudíž nejsou předmětem hodnocení. U obr. 13 i v textu chybí vysvětlení, co znamenají bodová data z července a srpna 2015 (místa sběru spektrometrických dat?). V části popisující předzpracování obrazových dat se uvádí, že byla vynechána pásma korespondující absorpčnímu pásmu vody. Co pak představují hodnoty na ose Y („data value“), která na první pohled odpovídají hodnotám spektrální odrazivosti vegetace, ovšem s hodnotami 1 v absorpčních pásmech vody? Na str. 40 má být uvedeno „Na obrázku 18“ (nikoli „XX“). Výhody a nevýhody metody DCA bych doporučovala uvést v rešeršní části práce, nikoli v metodice. Pro názornost bych přidala ukázkou

výřezu tabulky vstupující do programu Canoco 5. Výběr proměnných pro metodu PLSR, analýza hlavních komponent i metodika vlastní klasifikace jsou popsány dostatečně pro pochopení dále uvedených výsledků. Objevují se opět neobratná vyjádření jako např. „Výsledné rovnice získané z analýzy PLSR pro 3 typy spektrálních pásem ...“, spíše 3 typy dat? Obecně zvolený metodický postup hodnotím jako vhodný pro daný typ dat. Náročnost pochopení použitých teoretických modelů i komplexnost použitých dat odpovídá požadavkům na úroveň diplomové práce. Výběr je podepřen dostatečným množstvím příkladů z literatury.

Kapitola výsledky přehledně shrnuje výstupy DCA, PLSR i vlastní klasifikace. Výsledky Tab. 2 jsou interpretovány tak, že relativně nízké hodnoty vysvětlené variability v datech je dán velkým množstvím druhů a snímků. Lze tento výsledek chápat i tak, že podobnost jednotlivých fytoecologických snímků je nízká, resp. zastoupení jednotlivých druhů v těchto snímcích je jedinečné? V Tab. 3 a obr. 29 – 31 chybí vysvětlení třídy 8. Jaký je význam uvedeného počtu cifer (6 desetinných míst?) v Tab. 3? Jak velký podíl celkové informace byl vysvětlen použitými 10 pásmy PCA? Vzhledem k tomu, že výsledek klasifikace je hodnocen pouze na základě vizuální interpretace, jistě by stálo za zamyšlení jiného grafického vyjádření při porovnání barevné hodnoty z gradientového scatter plotu a klasifikovaného obrazu. Z uvedené vizualizace je hodnocení přesnosti velmi obtížné.

Diskuse je z hlediska porovnání s publikovanými pracemi podobného zaměření kvalitně zpracovaná. U vlastních výstupů bych kladla větší důraz na vysvětlení dosažených výsledků a hledání možných příčin než pouhé uvedení cifer a jejich porovnání. Nelze než souhlasit s tvrzením uvedeným v závěru, že použité metody mají svůj potenciál při studiu vazeb mezi druhovým složením zejména heterogenní vegetace a daty dálkového průzkumu a že by měly být předmětem dalšího výzkumu.

Otázky do diskuse:

1. Jak si vysvětlujete vyšší hodnotu koeficientu determinace  $R^2$  pro 1. osu DCA v případě spektrálních dat z terénu oproti obrazovým datům?
2. Jak vysvětlíte výrazné rozdíly ve výsledku klasifikace s využitím původních spekter obrazových dat a s využitím hlavních komponent?
3. Bylo by možné přesnost klasifikace, popř. porovnání výsledku jednotlivých výstupů, nějakým způsobem kvantifikovat a pak dále vizualizovat? Neuvažoval jste např. o vhodné metrice v RGB prostoru?
4. Co je příčinou paprskového vzoru v klasifikovaných datech (obr. 33 – 38)?
5. Kde vidíte výhody použitého přístupu vůči subpixelové klasifikaci obrazových dat?

I přes uvedené připomínky považuji předloženou diplomovou práci za kvalitně zpracovanou. Doporučuji ji k obhajobě a navrhuji hodnocení **velmi dobře**.